

5	Data Reduction	2
5.1	Sufficient Statistics	2
5.2	Factorization Theorem	5
5.3	Rao-Blackwell	10
5.4	Completeness	13
5.5	Exponential Class	17
5.6	Lehmann-Scheffe	20
5.7	Basu Theorem	25

5.1 Sufficient Statistics

Definition 1. Jointly Sufficient Statistics

Let $\mathbf{X} = (X_1, \dots, X_n)$ have joint pdf $f(\mathbf{x}, \boldsymbol{\theta})$, and let $S = (S_1, \dots, S_k)$ be a k -dimensional statistic. Then S_1, \dots, S_k is a set of **jointly sufficient statistics** for $\boldsymbol{\theta}$ if for any other vector of statistics, \mathbf{T} , the conditional pdf of \mathbf{T} given $\mathbf{S} = \mathbf{s}$, denoted by $f_{\mathbf{T}|\mathbf{S}}(t)$, does not depend on $\boldsymbol{\theta}$. In the one-dimensional case, we simply say that S is a **sufficient statistic** for θ .

Example 1.

Let X_1, \dots, X_n be a random sample from a Bernoulli distribution. Show that $S = \sum X_i$ is a sufficient statistic for θ by definition.

Example 2.

Consider a random sample from an exponential distribution, $X_i \sim EXP(\theta)$. Show that $S = \sum X_i$ is a sufficient statistic for θ by definition.

<div data-bbox="73 10 527 52" data-label="Section-Header"> <h2>5.2 Factorization Theorem</h2> </div> <div data-bbox="73 73 641 115" data-label="Section-Header"> <h3>Theorem 1. Factorization Criterion</h3> </div> <div data-bbox="73 115 714 231" data-label="Text"> <p>If X_1, \dots, X_n have joint pdf $f(x_1, \dots, x_n; \theta)$, and if $S = (S_1, \dots, S_k)$, then S_1, \dots, S_k are jointly sufficient for θ if and only if</p> </div> <div data-bbox="129 241 657 289" data-label="Equation-Block"> $f(x_1, \dots, x_k; \theta) = g(\mathbf{s}; \theta)h(x_1, \dots, x_n)$ </div> <div data-bbox="73 304 714 420" data-label="Text"> <p>where $g(\mathbf{s}; \theta)$ does not depend on x_1, \dots, x_n, except through s, and $h(x_1, \dots, x_n)$ does not involve θ.</p> </div> <div data-bbox="73 892 714 919" data-label="Page-Footer"> <div>MEME15203 STATISTICAL INFERENCE©DR YONG CHIN KHIAN</div> <div>202401</div> </div>	<div data-bbox="860 10 1047 52" data-label="Section-Header"> <h3>Example 3.</h3> </div> <div data-bbox="860 52 1502 220" data-label="Text"> <p>Consider a random sample from a Gamma distribution, $X_i \sim \text{Gamma}(\alpha = 3, \theta)$. Show that $S = \sum X_i$ is a sufficient statistic for θ by factorization theorem.</p> </div> <div data-bbox="860 892 1502 919" data-label="Page-Footer"> <div>MEME15203 STATISTICAL INFERENCE©DR YONG CHIN KHIAN</div> <div>202401</div> </div>
---	--

Definition 2. If A is a set, then the indicator function of A , denoted by I_A is defined as

$$I_A = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}$$

Example 4. Consider a random sample from a uniform distribution, $X_i \sim U(0, \theta)$, where θ is unknown. Find a sufficient statistic for θ .

Example 5. Consider a random sample from a uniform distribution, $X_i \sim U(\theta, \theta + 1)$. Notice that the length of the interval is one unit, but the endpoints are assumed to be unknown. Show that $S_1 = X_{1:n}$ and $S_2 = X_{n:n}$ are jointly sufficient for θ .

Example 6.

Consider a random sample from a normal distribution, $X_i \sim N(\mu, \sigma^2)$, where both μ and σ^2 are unknown. Show that $S_1 = \sum X_i$ and $S_2 = \sum X_i^2$ are jointly sufficient for $\theta = (\mu, \sigma^2)$.

5.3 Rao-Blackwell

Theorem 2. Rao-Blackwell Let X_1, \dots, X_n , have joint pdf $f(x_1, \dots, x_n; \theta)$, and let $S = (S_1, \dots, S_k)$ be a vector of jointly sufficient statistics for θ . If T is any unbiased estimator of $\tau(\theta)$, and if $T^* = E(T|S)$, then

1. T^* is an unbiased estimator of $\tau(\theta)$
2. T^* is a function of S , and
3. $V(T^*) \leq V(T)$ for every θ , and $V(T^*) < V(T)$ for some θ unless $T^* = T$ with probability 1.

Example 7.

Let X_1, X_2, \dots, X_n be random sample of size n from a Poisson distribution with unknown mean λ . Let

$$T = \begin{cases} 1, & X_1 = x \\ 0, & \text{otherwise} \end{cases}.$$

Find $E \left[T \mid \sum_{i=1}^n X_i = s \right]$.

Example 8.

Let X_1, X_2, \dots, X_n be random sample of size n from an Exponential distribution with unknown mean θ . Find the UMVUE of $\gamma = 1 - e^{-t/\theta}$ using Rao-Blackwell theorem.

5.4 Completeness

Definition 3. Completeness A family of density functions $\{f_{\mathbf{T}}(t; \boldsymbol{\theta}); \boldsymbol{\theta} \in \Omega\}$ is called complete if $E[u(T)] = 0$ for all $\boldsymbol{\theta} \in \Omega$ implies $u(T) = 0$ with probability 1 for all $\boldsymbol{\theta} \in \Omega$.

A sufficient statistic the density of which is a member of a complete family of density functions will be referred to as a **complete sufficient statistic**.

Example 9.

Let X_1, \dots, X_n denote a random sample from a Geometric distribution, $X_i \sim \textit{Geometric}(p)$. Show that $S = \sum X_i$ is the complete sufficient statistic for p .

Example 10.

Let X_1, \dots, X_n be iid $N(\theta, a\theta^2)$, where a is known constant and $\theta > 0$. Show that the family of distribution is not complete.

Example 11.

Consider a random sample of size n from a uniform distribution $X \sim U(0, \theta)$. Show that $S = X_{n:n}$ is a complete sufficient statistic for θ .

5.5 Exponential Class

Definition 4. Exponential Class A density function is said to be a member of the regular exponential class if it can be expressed in the form

$$f(\mathbf{x}; \boldsymbol{\theta}) = c(\boldsymbol{\theta})h(\mathbf{x})e^{-\sum_{j=1}^n q_j(\boldsymbol{\theta})t_j(\mathbf{x})}, \mathbf{x} \in A$$

and zero otherwise, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ is a vector of k unknown parameters, if the parameter space has the form

$$\Omega = \{\boldsymbol{\theta} | a_i \leq \theta_i \leq b_i, i = 1, \dots, k\}$$

(note that $a_i = -\infty$ and $b_i = \infty$ are permissible values), and if it satisfies regularity conditions 1, 2, and 3(a) or 3(b) given by:

- 1. The set $A = \{x : f(x; \boldsymbol{\theta}) > 0\}$ does not depend on $\boldsymbol{\theta}$.
- 2. The functions $q_j(\theta)$ are nontrivial, functionally independent, continuous functions of the θ .
- 3.(a) For a continuous random variable, the derivatives $t_j(x)$ are linearly independent contin-

uous functions of x over A .

- (b) For a discrete random variable, the $t_j(x)$ are nontrivial functions of x on A , and none is a linear function of the others.

For convenience, we will write that $f(x; \boldsymbol{\theta})$ is a member of $REC(q_1, \dots, q_k)$ or simply REC .

Theorem 3. X_1, \dots, X_n , is a random sample from a member of the regular exponential class $REC(q_1, \dots, q_k)$, then the statistics

$$S_1 = \sum_{i=1}^n t_1(X_i), \dots, S_k = \sum_{i=1}^n t_k(X_i)$$

are a minimal set of complete sufficient statistics for $\theta_1, \dots, \theta_k$.

Example 12.

Show that $X \sim Gamma(\alpha, \theta)$ belong to the regular exponential class, and use this information to find complete sufficient statistics based on a random sample X_1, \dots, X_n .

5.6 Lehmann-Scheffe

Theorem 4. Lehmann-Scheffe Let X_1, \dots, X_n have joint pdf $f(x_1, \dots, x_n; \boldsymbol{\theta})$, and let S be a vector of jointly complete sufficient statistics for $\boldsymbol{\theta}$. If $T^* = t^*(S)$ is a statistic that is unbiased for $\tau(\boldsymbol{\theta})$ and a function of S , then T^* is a UMVUE of $\tau(\boldsymbol{\theta})$.

<p>Example 13.</p> <p>Let X_1, \dots, X_n be a random sample from a Bernoulli distribution, $X_i \sim \text{Bin}(1, p)$; $0 < p < 1$. Find the UMVUE of p^2.</p>	<p>Example 14.</p> <p>Let X_1, \dots, X_{30} be a random sample from a distribution with probability density function(p.d.f.)</p> $f(x) = \frac{\theta^6}{\Gamma(6)} x^5 e^{-\theta x} I(0, \infty), \theta > 0.$ <p>(a) Show that the p.d.f. of X belongs to the regular exponential family.</p> <p>(b) Find a complete and sufficient statistic for θ.</p> <p>(c) Find the UMVUE for $V(X_1)$.</p> <p>(d) Find the UMVUE for θ.</p>
<p>MEME15203 STATISTICAL INFERENCE©DR YONG CHIN KHIAN</p> <p>202401</p>	<p>MEME15203 STATISTICAL INFERENCE©DR YONG CHIN KHIAN</p> <p>202401</p>

<div> <div>Chapter 5 Data Reduction23</div> <div> <p>Example 15.</p> <p>Let X_1, X_2, \dots, X_n be random sample of size n from $f(x \theta) = \binom{18}{x} \theta^x (1-\theta)^{18-x}, x = 0, 1, \dots, 18$. Find the uniformly minimum variance unbiased estimator(UMVUE) of $g(\theta) = \binom{18}{2} \theta^2 (1-\theta)^{16}$.</p> </div> </div>	<div> <div>Chapter 5 Data Reduction24</div> <div> <p>Example 16.</p> <p>Let X_1, X_2, \dots, X_n be a random sample from a distribution with pdf</p> $f(x; \theta) = 4\theta x^{4\theta-1}.$ <p>Find the UMVUE of θ,</p> </div> </div>
<p>MEME15203 STATISTICAL INFERENCE©DR YONG CHIN KHIAN</p> <p>202401</p>	<p>MEME15203 STATISTICAL INFERENCE©DR YONG CHIN KHIAN</p> <p>202401</p>

<div data-bbox="73 14 394 46" data-label="Section-Header"> <h3>5.7 Basu Theorem</h3> </div> <div data-bbox="73 71 716 386" data-label="Text"> <p>Theorem 5. Basu Let X_1, \dots, X_n, have joint pdf $f(x_1, \dots, x_n; \theta)$; $\theta \in \Omega$. Suppose that $S = (S_1, \dots, S_k)$ where S_1, \dots, S_k are jointly complete sufficient statistics for θ, and suppose that T is any other statistic. If the distribution of T does not involve θ, then S and T are stochastically independent. In this case, T is call an ancillary statistic.</p> </div> <div data-bbox="73 894 716 915" data-label="Page-Footer"> <div>MEME15203 STATISTICAL INFERENCE©DR YONG CHIN KHIAN</div> <div>202401</div> </div>	<div data-bbox="862 14 1503 168" data-label="Text"> <p>Example 17. Consider a random sample of size n from a normal distribution $X \sim N(\mu, \sigma^2)$. Use Basu Theorem to show that \bar{X} and S^2 are independent.</p> </div> <div data-bbox="862 894 1503 915" data-label="Page-Footer"> <div>MEME15203 STATISTICAL INFERENCE©DR YONG CHIN KHIAN</div> <div>202401</div> </div>
--	---

<div data-bbox="243 1068 716 1098" data-label="Page-Header"> <div>Chapter 5 Data Reduction</div> <div>27</div> </div> <div data-bbox="73 1127 781 1558" data-label="Text"> <p>Example 18.</p> <p>Consider a random sample of size n from a two-parameter exponential distribution, $X_i \sim EXP(1, \eta)$.</p> <p>(a) Show that $T(X) = X_{1:n}$ is complete and sufficient for η.</p> <p>(b) Find the UMVUE of η.</p> <p>(c) Find the UMVUE of η^2.</p> <p>(d) Use Basu’s Theorem to show that $X_{1:n}$ and $S^2 = \frac{\sum(X_i-\bar{X})^2}{n-1}$ are independent.</p> </div> <div data-bbox="73 2007 716 2028" data-label="Page-Footer"> <div>MEME15203 STATISTICAL INFERENCE©DR YONG CHIN KHIAN</div> <div>202401</div> </div>	
---	--